**Proposal for mapping of clinical documents to Clinical Document Ontology (CDO) and standard terminology**

HL7/LOINC CDO is a standard for consistent naming of documents to support a range of use cases: retrieval, organization, display, and exchange. It guides the creation of LOINC codes for clinical notes. CDO annotates each document with 5 dimensions:

* Kind of Document

Characterizes the general structure of the document at a macro level (e.g. Anesthesia Consent)

* Type of Service

Characterizes the kind of service or activity (e.g. evaluations, consultations, and summaries). The notion of time sequence, e.g., at the beginning (admission) at the end (discharge) is subsumed in this axis. Example: Discharge Teaching.

* Setting

Setting is an extension of CMS’s definitions (e.g. Inpatient, Outpatient)

* Subject Matter Domain (SMD)

Characterizes the subject matter domain of a note (e.g. Anesthesiology)

* Role

Characterizes the training or professional level of the author of the document, but does not break down to specialty or subspecialty (e.g. Physician)

Each combination of these 5 dimensions should roll up to a unique LOINC code. For example, Dentistry Hygienist Outpatient Progress note (LOINC code 34127-1) has the following dimensions:



According to CDO requirements, only 2 of the 5 dimensions are required to properly annotate a document: Kind of Document and any one of the other 4 dimensions.

However, not all the permutations of the CDO dimensions will necessarily yield an existing LOINC code.2 HL7/LOINC workforce is committed to establish new LOINC codes for each new encountered combination of CDO dimensions. 3

Automation of mapping of clinical notes to a standard terminology based on the note title is possible when it is driven by ontology (aka CDO). Mapping to individual LOINC codes which may or may not exist for a particular note type cannot be fully automated. To support mapping of clinical notes to CDO in OMOP CDM, we propose the following approach.

1. Add all LOINC concepts representing 5 CDO dimensions to the Concept table. For example:

|  |  |  |
| --- | --- | --- |
| **Field** | **Record 1** | **Record 2** |
| concept\_id | 55443322132 | 55443322175 |
| concept\_name | Administrative note | Against medical advice note |
| concept\_code | LP173418-7 | LP173388-2 |
| vocabulary\_id | LOINC | LOINC |

1. Represent CDO hierarchy in the Concept\_Relationship table using the “*Subsumes*” – “*Is a*” relationship pair. For example:

|  |  |  |
| --- | --- | --- |
| **Field** | **Record 1** | **Record 2** |
| concept\_id\_1 | 55443322132 | 55443322175 |
| concept\_id\_2 | 55443322175 | 55443322132 |
| relationship\_id | *Subsumes* | *Is a* |

“LP173387-4 Administrative note” “*Subsumes*” “LP173388-2 Against medical advice note”

“LP173388-2 Against medical advice note” “*Is a*” “LP173387-4 Administrative note”

These hierarchies should also be represented in the Concept\_Ancestor table.

1. Add LOINC document codes to the Concept table (e.g. Dentistry Hygienist Outpatient Progress note, LOINC code 34127-1). For example:

|  |  |  |
| --- | --- | --- |
| **Field** | **Record 1** | **Record 2** |
| concept\_id | 193240 | 193241 |
| concept\_name | Dentistry Hygienist Outpatient Progress note | Consult note |
| concept\_code | 34127-1 | 11488-4 |
| vocabulary\_id | LOINC | LOINC |

1. Represent dimensions of each document concept in Concept\_Relationship table by its relationships to the respective concepts from CDO. Use the “Member Of” – “Has Member” (new) relationship pair. Using example from the Dentistry Hygienist Outpatient Progress note (LOINC code 34127-1):

|  |  |  |
| --- | --- | --- |
| **concept\_id\_1** | **concept\_id\_1** | **relationship\_id** |
| 193240 | 55443322132 | Member Of |
| 55443322132 | 193240 | Has Member |
| 193240 | 55443322175 | Member Of |
| 55443322175 | 193240 | Has Member |
| 193240 | 55443322166 | Member Of |
| 55443322166 | 193240 | Has Member |
| 193240 | 55443322107 | Member Of |
| 55443322107 | 193240 | Has Member |
| 193240 | 55443322146 | Member Of |
| 55443322146 | 193240 | Has Member |

Where concept codes represent the following concepts:

|  |  |
| --- | --- |
| **Content** | **Description** |
| 193240 | Corresponds to LOINC 34127-1, Dentistry Hygienist Outpatient Progress note |
| 55443322132 | Corresponds to LOINC LP173418-7, Kind of Document = Note |
| 55443322175 | Corresponds to LOINC LP173213-2, Type of Service = Progress |
| 55443322166 | Corresponds to LOINC LP173051-6, Setting = Outpatient |
| 55443322107 | Corresponds to LOINC LP172934-4, Subject Matter Domain = Dentistry |
| 55443322146 | Corresponds to LOINC LP173071-4, Role = Hygienist |

Most of the codes will not have all 5 dimensions. Therefore, they may be represented by 2-5 relationship pairs.

1. If LOINC does not have a code corresponding to a permutation of the 5 CDO encountered in the source, this code will be generated as OMOP vocabulary code. Its relationships to the CDO dimensions will be represented exactly as those of existing LOINC concepts (as described above). If/when a proper LOINC code for this permutation is released, the old code should be deprecated. Transition between the old and new codes should be represented by “Concept replaces” – “Concept replaced by” pairs.
2. Mapping from the source data will be performed to the 2-5 CDO dimensions.

Query below finds LOINC code for Dentistry Hygienist Outpatient Progress note (see example above) that has all 5 dimensions:

SELECT

FROM Concept\_Relationship

WHERE relationship\_id = ‘Has Member’ AND

(concept\_id\_1 = 55443322132

OR concept\_id\_1 = 55443322175

OR concept\_id\_1 = 55443322166

OR concept\_id\_1 = 55443322107

OR concept\_id\_1 = 55443322146)

GROUP BY concept\_ID\_2

If less than 5 dimensions are available, HAVING COUNT(n) clause should be added to get a unique record at the intersection of these dimensions. n is the number of dimensions available:

SELECT

FROM Concept\_Relationship

WHERE relationship\_id = ‘Has Member’ AND

(concept\_id\_1 = 55443322132

OR concept\_id\_1 = 55443322175

OR concept\_id\_1 = 55443322146)

GROUP BY concept\_ID\_2

HAVING COUNT(\*) = 3

The proposed approach will ensure that any combination of the 5 CDO dimensions encountered in the source data has a corresponding concept in the vocabulary. It will also support consistent approach to the OMOP CDM/Vocabulary conventions:

* One required \_type\_concept\_id field will be populated in a corresponding domain table, NOTE.
* Vocabulary-related attributes are stored in a vocabulary data model in a uniform way
* Usage of a standard vocabulary, LOINC, is ensured where possible
* Introduction of new OMOP concepts when a standard does not provide adequate coverage of the source data

A similar mapping approach can be applied to labs.